max planck institut informatik

Author-Specific Sentiment Aggregation for Rating Prediction of Reviews

Subhabrata Mukherjee [1] **Sachindra Joshi** [2]

[1] **Max Planck Institute for Informatics**

[2] **IBM India Research Lab**

***L*REC 2014**

May 31, 2014

Outline

Objective

Classify a piece of text as positive, negative or objective

I saw a movie 

The direction was awesome 

However, the acting was not that good. 

IMDB Review

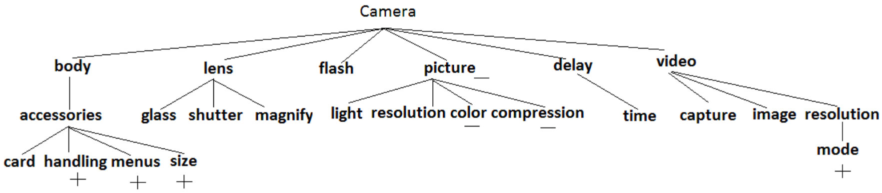
"I bought a Canon EOS 7D (DSLR). It's very small, sturdy, and constructed well. The handling is quite nice with a powder-coated metal frame. It powers on quickly and the menus are fairly easy to navigate. The video modes are nice, too. It works great with my 8GB Eye-Fi SD card. A new camera isn't worth it if it doesn't exceed the picture quality of my old 5Mpixel SD400 and this one doesn't. The auto white balance is poor. I'd need to properly balance every picture taken so far with the ELPH 300. With 12 Mpixels, you'd expect pretty good images, but the problem is that the ELPH 300 compression is turned up so high that the sensor's acuity gets lost (softened) in compression."

camera size, structure, easy use, video modes, SD support

auto-white balance, high compression leading to sensor acuity

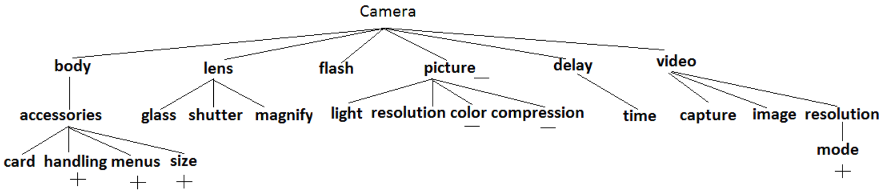
overall review polarity

Camera Sentiment Ontology Tree



- Ontology is a knowledge base of structured list of concepts, relations and individuals
- Hierarchical relationship between product attributes are captured
- Overall polarity negative as facet polarities higher up the tree *dominate* those at a lower level

Camera Sentiment Ontology Tree



- Ontology is a knowledge base of structured list of concepts, relations and individuals
- Hierarchical relationship between product attributes are captured
- Overall polarity negative as facet polarities higher up the tree *dominate* those at a lower level

Why Author Specificity ?

"[This film is based on a true-life incident. It sounds like a great plot and the director makes a decent attempt in narrating a powerful story.] [However, the film does not quite make the mark due to sloppy acting.]"

- Rating varies for authors with different topic preferences
  - Positive for those with preference for *acting* and *narration*
  - Negative for *acting*
- Affective sentiment varies for authors
  - How much negative is "does not quite make the mark" for me ?
- Author-writing style helps associate facets and sentiments[1]
  - E.g. topic switch, use of content and function words etc.
  - The author makes a topic switch with the function word "however"
- Traditional works ignore *author* identity

[1] Joint Author Sentiment Topic Model, Mukherjee S. et al., Siam Data Mining (SDM) 2014

Why Author Specificity ?

“[This film is based on a true-life incident. It sounds like a great plot and the director makes a decent attempt in narrating a powerful story.] [However, the film does not quite make the mark due to sloppy acting.]”

- Rating varies for authors with different topic preferences
  - Positive for those with preference for *acting* and *narration*
  - Negative for *acting*
- Affective sentiment varies for authors
  - How much negative is “does not quite make the mark” for me ?
- Author-writing style helps associate facets and sentiments[1]
  - E.g. topic switch, use of content and function words etc.
  - The author makes a topic switch with the function word “however”
- Traditional works ignore *author* identity

[1] Joint Author Sentiment Topic Model, Mukherjee S. et al., Siam Data Mining (SDM) 2014

Why Author Specificity ?

“[This film is based on a true-life incident. It sounds like a great plot and the director makes a decent attempt in narrating a powerful story.] [However, the film does not quite make the mark due to sloppy acting.]”

- Rating varies for authors with different topic preferences
  - Positive for those with preference for *acting* and *narration*
  - Negative for *acting*
- Affective sentiment varies for authors
  - How much negative is “does not quite make the mark” for me ?
- Author-writing style helps associate facets and sentiments[1]
  - E.g. topic switch, use of content and function words etc.
  - The author makes a topic switch with the function word “however”
- Traditional works ignore *author* identity

[1] Joint Author Sentiment Topic Model, Mukherjee S. et al., Siam Data Mining (SDM) 2014

Why Author Specificity ?

"[This film is based on a true-life incident. It sounds like a great plot and the director makes a decent attempt in narrating a powerful story.] [However, the film does not quite make the mark due to sloppy acting.]"

- Rating varies for authors with different topic preferences
  - Positive for those with preference for *acting* and *narration*
  - Negative for *acting*
- Affective sentiment varies for authors
  - How much negative is "does not quite make the mark" for me ?
- Author-writing style helps associate facets and sentiments[1]
  - E.g. topic switch, use of content and function words etc.
  - The author makes a topic switch with the function word "however"
- Traditional works ignore *author* identity

[1] Joint Author Sentiment Topic Model, Mukherjee S. et al., Siam Data Mining (SDM) 2014

Why Author Specificity ?

"[This film is based on a true-life incident. It sounds like a great plot and the director makes a decent attempt in narrating a powerful story.] [However, the film does not quite make the mark due to sloppy acting.]"

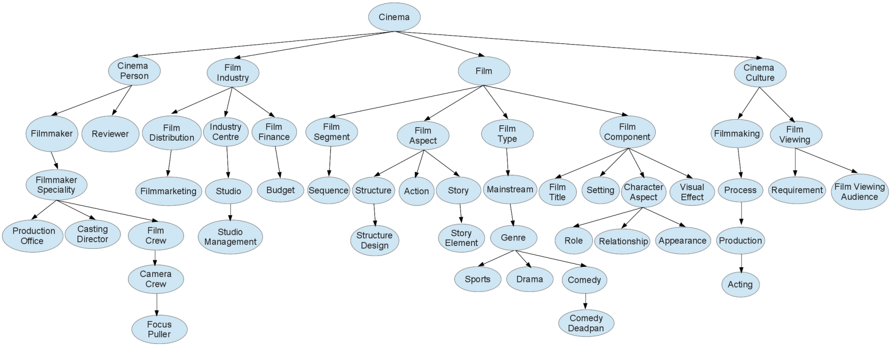
- Rating varies for authors with different topic preferences
  - Positive for those with preference for *acting* and *narration*
  - Negative for *acting*
- Affective sentiment varies for authors
  - How much negative is "does not quite make the mark" for me ?
- Author-writing style helps associate facets and sentiments[1]
  - E.g. topic switch, use of content and function words etc.
  - The author makes a topic switch with the function word "however"
- Traditional works ignore *author* identity

[1] Joint Author Sentiment Topic Model, Mukherjee S. et al., Siam Data Mining (SDM) 2014

Review

"as with any gen-x mtv movie (like last year's dead man on campus), the movie is marketed for a primarily male audience as indicated by its main selling points: sex and football. those two items are sure to snare a sizeable box office chunk initially, but sales will decline for two reasons. first, the football sequences are nothing new; the sports genre isn't mainstream and it's been retread to death. second, the sex is just bad. despite the appearance of a whipped cream bikini or the all-night strip-club party, there's nothing even remotely tantalizing. the acting is mostly mediocre, not including the fantastic jon voight. cultivating his usual sliminess, voight gives an unexpectedly standout performance as west canaan coyotes head coach bud kilmer ... these elements (as well as the heavy drinking and carousing) might be more appropriate on a college campus - but mtv's core audience is the high school demographic. this focus is further emphasized by the casting: james van der beek, of tv's "dawson's creek", is an understandable choice for the reluctant hero..."

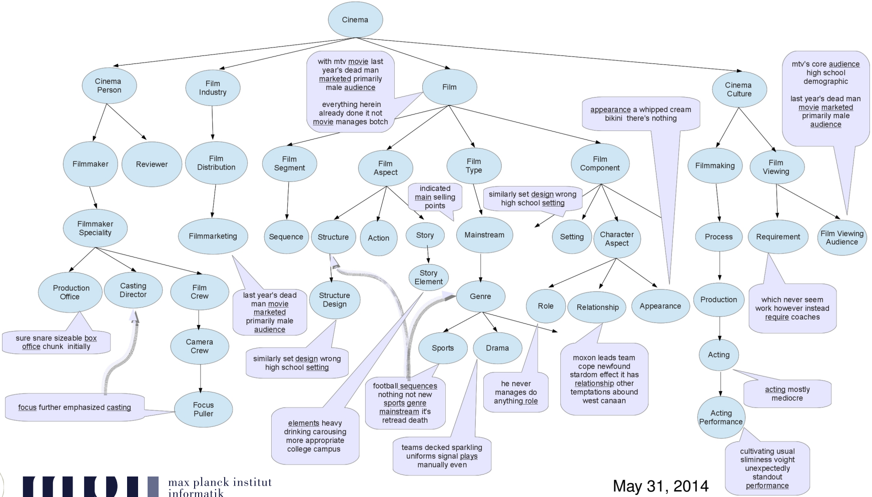
Cinema Ontology Tree



JedFilm. (2014). Cinema ontology project, March.

Phrase Annotated Sentiment Ontology Tree

- Concept mapping from review to SOT using Wu-Palmer WordNet similarity measure
- Facet-Specific Opinion Extraction using Dependency Parsing (Mukherjee et al., CICLING 2012)



Phrase Annotated Author-Specific SOT

$T(V, E)$	SOT
$V_j =< f_j, < p_i^j >, w_j, d_j >$	Product Attribute Set
$E_{j,k}$	Attribute relation connecting V_j and V_k
f_j	Product Facet
p_i^j	<Phrases in the review with author opinion about f_j>
w_j	Author preference about f_j
d_j	Depth of f_j in SOT
$O(p)$	Sentiment Predictor Function that maps $polarity(p_i^j) \in [-1, 1]$
PASOT	Equipped with $(T^a(V, E), O^a(p))$ for a given author *a*

Expected Sentiment Weight

Expected sentiment weight (ESW) of a node in *PASOT* is defined as,

$$ESW^a(V_j) = \underbrace{\overbrace{w_j^a}^{\text{author-preference}} \times \overbrace{\frac{1}{d_j}}^{\text{facet depth}} \times \sum_i \overbrace{O^a(p_i^j)}^{\text{phrase polarity}}}_{\text{self-information}} + \underbrace{\sum_k ESW^a(V_{j,k})}_{\text{children information}} \quad (1)$$

where $O^a(p_i^j) \in [-1, 1]$

Review polarity given by $ESW^a(root)$

Computation of ESW requires learning $< w_j^a >$ and O^a for each author a

Review Polarity

- For each author a, every facet f_j is associated with $ESW^a(V_j)$, where $f_j \in V_j$, computed using Equation 1
- Let y_i be the overall polarity of review i
- To learn overall review polarity as a function of author-specific facet preferences

Formulate an L_2-regularized logistic regression problem :

$$\min_{w^a} \frac{1}{2} {w^a}^T w^a + C \sum_i \log(1 + \exp^{-y_i \sum_j w_j^a \times ESW^a(V_j)}) \tag{2}$$

Trust region newton method (Lin et al., JMLR 2008) used to learn the weights in the above equation

Review Polarity

- For each author a, every facet f_j is associated with $ESW^a(V_j)$, where $f_j \in V_j$, computed using Equation 1
- Let y_i be the overall polarity of review i
- To learn overall review polarity as a function of author-specific facet preferences

Formulate an L_2-regularized logistic regression problem :

$$\min_{w^a} \frac{1}{2} {w^a}^T w^a + C \sum_i \log(1 + \exp^{-y_i \sum_j w_j^a \times ESW^a(V_j)}) \qquad (2)$$

Trust region newton method (Lin et al., JMLR 2008) used to learn the weights in the above equation

Review Polarity

- For each author a, every facet f_j is associated with $ESW^a(V_j)$, where $f_j \in V_j$, computed using Equation 1
- Let y_i be the overall polarity of review i
- To learn overall review polarity as a function of author-specific facet preferences

Formulate an L_2-regularized logistic regression problem :

$$\min_{w^a} \frac{1}{2} {w^a}^T w^a + C \sum_i \log(1 + \exp^{-y_i \sum_j w_j^a \times ESW^a(V_j)}) \tag{2}$$

Trust region newton method (Lin et al., JMLR 2008) used to learn the weights in the above equation

Review Polarity

- For each author a, every facet f_j is associated with $ESW^a(V_j)$, where $f_j \in V_j$, computed using Equation 1
- Let y_i be the overall polarity of review i
- To learn overall review polarity as a function of author-specific facet preferences

Formulate an L_2-regularized logistic regression problem :

$$\min_{w^a} \frac{1}{2} {w^a}^T w^a + C \sum_i \log(1 + \exp^{-y_i \sum_j w_j^a \times ESW^a(V_j)}) \qquad (2)$$

Trust region newton method (Lin et al., JMLR 2008) used to learn the weights in the above equation

Sentiment Predictor Function $O(p)$

L_2-regularized L_2-loss Support Vector Machine and bag-of-words unigram features trained over the movie review corpus in (Maas et al., ACL 2011).

Author-Specific Hierarchical Sentiment Aggregation

- Learn domain-specific ontology $T(V, E)$ using KB
- Learn phrase polarity predictor $O(p)$ using L_2-reg. L_2-loss SVM
- For each author
  - Map each review to $T(V, E)$ using Wu-Palmer Similarity
  - Use Dependency Parsing Algorithm (Mukherjee S. et al., CICLING 2012) to extract feature-specific opinion $< p_j^i >$
  - Construct PASOT for each review using $O(p_j^i)$
  - Apply Eqn 1 to PASOT bottom-up to find $ESW(V_j)\ \forall j$
  - Using 80% labeled reviews y_i and $< ESW^a(V_j) >$, learn author-specific facet-weights $< w_j^a >$ using Equation 2
  - For each unseen review
    - Construct *PASOT* using above steps and learnt w^a
    - Use Equation 1 to find $< ESW^a(V_j) >$
    - Review polarity given by $Sign(ESW^a(root))$

Author-Specific Hierarchical Sentiment Aggregation

- Learn domain-specific ontology $T(V, E)$ using KB
- Learn phrase polarity predictor $O(p)$ using L_2-reg. L_2-loss SVM
- For each author
  - Map each review to $T(V, E)$ using Wu-Palmer Similarity
  - Use Dependency Parsing Algorithm (Mukherjee S. et al., CICLING 2012) to extract feature-specific opinion $< p_j^i >$
  - Construct PASOT for each review using $O(p_j^i)$
  - Apply Eqn 1 to PASOT bottom-up to find $ESW(V_j)\ \forall j$
  - Using 80% labeled reviews y_i and $< ESW^a(V_j) >$, learn author-specific facet-weights $< w_j^a >$ using Equation 2
  - For each unseen review
    - Construct *PASOT* using above steps and learnt w^a
    - Use Equation 1 to find $< ESW^a(V_j) >$
    - Review polarity given by $Sign(ESW^a(root))$

Author-Specific Hierarchical Sentiment Aggregation

- Learn domain-specific ontology $T(V, E)$ using KB
- Learn phrase polarity predictor $O(p)$ using L_2-reg. L_2-loss SVM
- For each author
  - Map each review to $T(V, E)$ using Wu-Palmer Similarity
  - Use Dependency Parsing Algorithm (Mukherjee S. et al., CICLING 2012) to extract feature-specific opinion $< p_j^i >$
  - Construct PASOT for each review using $O(p_j^i)$
  - Apply Eqn 1 to PASOT bottom-up to find $ESW(V_j)\ \forall j$
  - Using 80% labeled reviews y_i and $< ESW^a(V_j) >$, learn author-specific facet-weights $< w_j^a >$ using Equation 2
  - For each unseen review
    - Construct *PASOT* using above steps and learnt w^a
    - Use Equation 1 to find $< ESW^a(V_j) >$
    - Review polarity given by $Sign(ESW^a(root))$

Author-Specific Hierarchical Sentiment Aggregation

- Learn domain-specific ontology $T(V, E)$ using KB
- Learn phrase polarity predictor $O(p)$ using L_2-reg. L_2-loss SVM
- For each author
  - Map each review to $T(V, E)$ using Wu-Palmer Similarity
  - Use Dependency Parsing Algorithm (Mukherjee S. et al., CICLING 2012) to extract feature-specific opinion $< p_j^i >$
  - Construct PASOT for each review using $O(p_j^i)$
  - Apply Eqn 1 to PASOT bottom-up to find $ESW(V_j)\ \forall j$
  - Using 80% labeled reviews y_i and $< ESW^a(V_j) >$, learn author-specific facet-weights $< w_j^a >$ using Equation 2
  - For each unseen review
    - Construct *PASOT* using above steps and learnt w^a
    - Use Equation 1 to find $< ESW^a(V_j) >$
    - Review polarity given by $Sign(ESW^a(root))$

Author-Specific Hierarchical Sentiment Aggregation

- Learn domain-specific ontology $T(V, E)$ using KB
- Learn phrase polarity predictor $O(p)$ using L_2-reg. L_2-loss SVM
- For each author
  - Map each review to $T(V, E)$ using Wu-Palmer Similarity
  - Use Dependency Parsing Algorithm (Mukherjee S. et al., CICLING 2012) to extract feature-specific opinion $< p_j^i >$
  - Construct PASOT for each review using $O(p_j^i)$
  - Apply Eqn 1 to PASOT bottom-up to find $ESW(V_j)\ \forall j$
  - Using 80% labeled reviews y_i and $< ESW^a(V_j) >$, learn author-specific facet-weights $< w_j^a >$ using Equation 2
  - For each unseen review
    - Construct *PASOT* using above steps and learnt w^a
    - Use Equation 1 to find $< ESW^a(V_j) >$
    - Review polarity given by $Sign(ESW^a(root))$

Author-Specific Hierarchical Sentiment Aggregation

- Learn domain-specific ontology $T(V, E)$ using KB
- Learn phrase polarity predictor $O(p)$ using L_2-reg. L_2-loss SVM
- For each author
  - Map each review to $T(V, E)$ using Wu-Palmer Similarity
  - Use Dependency Parsing Algorithm (Mukherjee S. et al., CICLING 2012) to extract feature-specific opinion $< p_j^i >$
  - Construct PASOT for each review using $O(p_j^i)$
  - Apply Eqn 1 to PASOT bottom-up to find $ESW(V_j)\ \forall j$
  - Using 80% labeled reviews y_i and $< ESW^a(V_j) >$, learn author-specific facet-weights $< w_j^a >$ using Equation 2
  - For each unseen review
    - Construct *PASOT* using above steps and learnt w^a
    - Use Equation 1 to find $< ESW^a(V_j) >$
    - Review polarity given by $Sign(ESW^a(root))$

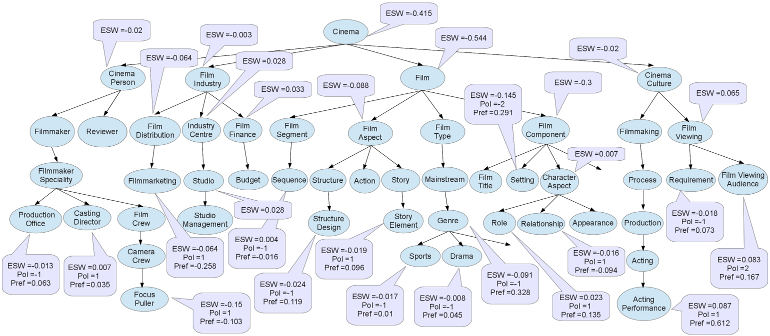
Author-Specific Hierarchical Sentiment Aggregation

- Learn domain-specific ontology $T(V, E)$ using KB
- Learn phrase polarity predictor $O(p)$ using L_2-reg. L_2-loss SVM
- For each author
  - Map each review to $T(V, E)$ using Wu-Palmer Similarity
  - Use Dependency Parsing Algorithm (Mukherjee S. et al., CICLING 2012) to extract feature-specific opinion $< p_j^i >$
  - Construct PASOT for each review using $O(p_j^i)$
  - Apply Eqn 1 to PASOT bottom-up to find $ESW(V_j)\ \forall j$
  - Using 80% labeled reviews y_i and $< ESW^a(V_j) >$, learn author-specific facet-weights $< w_j^a >$ using Equation 2
  - For each unseen review
    - Construct *PASOT* using above steps and learnt w^a
    - Use Equation 1 to find $< ESW^a(V_j) >$
    - Review polarity given by $Sign(ESW^a(root))$

Author-Specific Hierarchical Sentiment Aggregation

- Learn domain-specific ontology $T(V, E)$ using KB
- Learn phrase polarity predictor $O(p)$ using L_2-reg. L_2-loss SVM
- For each author
  - Map each review to $T(V, E)$ using Wu-Palmer Similarity
  - Use Dependency Parsing Algorithm (Mukherjee S. et al., CICLING 2012) to extract feature-specific opinion $< p_j^i >$
  - Construct PASOT for each review using $O(p_j^i)$
  - Apply Eqn 1 to PASOT bottom-up to find $ESW(V_j)\ \forall j$
  - Using 80% labeled reviews y_i and $< ESW^a(V_j) >$, learn author-specific facet-weights $< w_j^a >$ using Equation 2
  - For each unseen review
    - Construct *PASOT* using above steps and learnt w^a
    - Use Equation 1 to find $< ESW^a(V_j) >$
    - Review polarity given by $Sign(ESW^a(root))$

Learnt PASOT for Example Review



Dataset

Dataset	Authors	Avg Rev/ Author	Rev/ Rating			Avg Rev Length	Avg Words/ Rev
			Pos	**Neg**	**Total**		
Movie Review*	312	7	1000	1000	2000	32	746
			Pos	**Neg**	**Total**		
Movie Review⊥	65	23	705	762	1467	32	711

Table: Movie Review Dataset Statistics (* denotes the original data, ⊥ indicates processed data)

Baselines

- Support Vector Machines with L_2-loss, L_2-reg and unigram bag-of-words features (Pang and Lee, EMNLP 2002; Pang and Lee, ACL 2004; Mullen and Collier, EMNLP 2004)
- Author-specific facet preference using regression (Mukherjee et al., WWW 2013)
- Sentiment aggregation using ConceptNet ontology (Mukherjee and Joshi, IJCNLP 2013)

Accuracy Comparison with Baselines

Model	Author Acc.	Overall Acc.
Bag-of-words Support Vector Machine (Pang and Lee, EMNLP 2002; Pang and Lee, ACL 2004; Mullen and Collier, EMNLP 2004)	80.23	78.49
Author-Specific Analysis using Regression (Mukherjee et al., WWW 2013)	79.31	79.07
Ontological Sentiment Aggregation (Mukherjee and Joshi, IJCNLP 2013)	81.4	79.51
PASOT	**86.32**	**86.04**

Comparison of Existing Models with PASOT in the IMDB Dataset

Models	Acc.
Eigen Vector Clustering (Dasgupta et al., EMNLP 2009)	70.9
Semi Supervised, 40% doc. Label (Li et al., IJCNLP 2009)	73.5
LSM Unsupervised with prior info (Lin et al., CIKM 2009)	74.1
SO-CAL Full Lexicon (Taboada et al., Comp. Ling. 2011)	76.37
RAE Semi Supervised Recursive Auto Encoders with random word initialization (Dasgupta et al., EMNLP 2009)	76.8
WikiSent: Extractive Summarization with Wikipedia + Lexicon (Mukherjee et al., ECML-PKDD 2012)	76.85
Supervised Tree-CRF (Socher et al., EMNLP 2011)	77.3
RAE: Supervised Recursive Auto Encoders with 10% cross-validation (Socher et al., EMNLP 2011)	77.7
JST: Without Subjectivity Detection using LDA (Lin et al., CIKM 2009)	82.8
Supervised SVM (Pang et al., EMNLP 2002)	82.9
JST: With Subjectivity Detection (Lin et al., CIKM 2009)	84.6
PASOT	**86.04**
Supervised SVM (Kennedy et al., Comp. Intell. 2006)	86.2
Supervised Subjective MR, SVM (Pang et al., ACL 2004)	87.2
JAST: Joint Author Sentiment Topic Model (Mukherjee et al., SDM 2014)	87.69
Appraisal Group: Supervised (Whitelaw et al., CIKM 2005)	90.2

Use Case : Thwarting Detection

Example of thwarting from (Pang and Lee, EMNLP 2002) :

"This film sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up."

- Overall review polarity different from that of majority opinion words in review.

Dataset	Positive Thwarted	Negative Thwarted
1467	279	132

Model	Thwarting Acc.
Bag-of-words SVM	61.54
PASOT	73.07

Table: Thwarting Accuracy Comparison

Use Case : Thwarting Detection

Example of thwarting from (Pang and Lee, EMNLP 2002) :

"This film sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up."

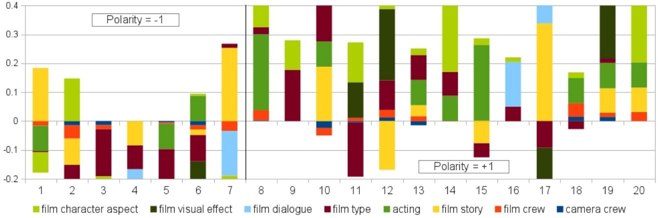
- Overall review polarity different from that of majority opinion words in review.

Dataset	Positive Thwarted	Negative Thwarted
1467	279	132

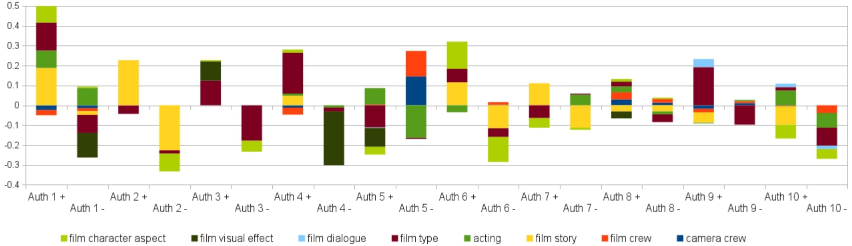
Model	Thwarting Acc.
Bag-of-words SVM	61.54
PASOT	73.07

Table: Thwarting Accuracy Comparison

Variation of ESW of Facets with Review Rating for a Specific Author



Variation of ESW of Facets with Review Rating for 10 Authors



Conclusions

- Hierarchical sentiment aggregation performs better than flat classification models
- Proposed an approach to construct a *Phrase Annotated Author-Specific Sentiment Ontology Tree* (PASOT)
- Author-specific modeling of reviews better capture author intention and facet preference leading to better prediction models

Questions ???